

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Рабочая программа дисциплины (модуля),
как компонент образовательной программы
специализированного высшего образования
по направлению подготовки
09.04.01 Информатика и вычислительная техника,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Инструменты анализа данных

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность (профиль): Искусственный интеллект и предиктивная аналитика в транспортных системах

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде электронного документа выгружена из единой корпоративной информационной системы управления университетом и соответствует оригиналу

Простая электронная подпись, выданная РУТ (МИИТ)
ID подписи: 5665
Подписал: заведующий кафедрой Нутович Вероника
Евгеньевна
Дата: 01.09.2026

1. Общие сведения о дисциплине (модуле).

Целью освоения данной дисциплины является получение базовых, теоретических знаний и навыков в области работы с данными и применения программных средств работы с данными для выполнения преобразований данных, анализа данных и визуализации результатов.

В рамках дисциплины у обучающихся формируются базовые представления и знания о жизненном цикле разработки моделей анализа данных и машинного обучения, лучших практиках автоматизации процессов анализа данных в соответствии с методологией MLOps.

Задачей дисциплины является формирование у обучающихся навыков работы с инструментами сбора и подготовки данных, контроля версий, оптимизации, обучения, развертывания и мониторинга моделей анализа данных и машинного обучения.

2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

ПК-2 - Способен осуществить сбор, очистку, подготовку и разметку данных используя методологию ETL для дальнейшего обучения моделей искусственного интеллекта;

ПК-3 - Способен проектировать, разрабатывать, обучать, оценивать и разворачивать модели искусственного интеллекта в соответствии с DevOps, DataOps и MLOps методологиями.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

Уметь:

- разрабатывать требования к эксплуатации моделей анализа данных и машинного обучения;

- использовать системы контроля версий для версионирования моделей и данных;

- использовать Apache Airflow для организации воспроизводимого конвейера машинного обучения;

- использовать средства мониторинга работы моделей анализа данных и машинного обучения.

Знать:

- особенности постановки бизнес-требований к анализу данных;

- основы документирования этапов процесса анализа данных;
- основные форматы моделей анализа данных и машинного обучения;
- основы работы с хранилищами моделей;
- основы работы в MLOps инструментах и платформах.

Владеть:

- навыками работы с инструментами сбора, хранения и подготовки данных в соответствии с концепцией ETL;
- навыками оптимизации гиперпараметров моделей анализа данных и машинного обучения;
- навыками работы с инструментами обучения моделей анализа данных и машинного обучения;
- навыками работы с фреймворком Flask и FastAPI для реализации развертываемых сервисов моделей;
- навыками интеграции моделей в текущих цифровых интеллектуальных системах.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 5 з.е. (180 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Семестр №1
Контактная работа при проведении учебных занятий (всего):	48	48
В том числе:		
Занятия лекционного типа	16	16
Занятия семинарского типа	32	32

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 132 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

№ п/п	Тематика лекционных занятий / краткое содержание
1	<p>Постановка задачи и анализ данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - введение в бизнес-требования к анализу данных, спецификациям и требованиям к эксплуатации разрабатываемой интеллектуальной системы; - документирование этапов процесса анализа данных, разработки и развертывания интеллектуальной системы; - жизненный цикл анализа данных и их связь с инструментами анализа данных.
2	<p>Сбор, хранение и подготовка данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - введение в понятие сбора, майнинга данных, очистка данных, балансировка данных; - концепция ETL; - банки и базы данных, дата-сет, открытые данные; - обзор инструментов сбора, очистки, балансировки и хранения данных; - введение в подготовку данных, разметку данных; - обзор инструментов разметки данных.
3	<p>Контроль версий.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - введение в проблематику воспроизводимости в анализе данных; - версионирование как данных, так и моделей, принципы и подходы; - обзор инструментов контроля версий для данных и моделей.
4	<p>Оптимизация гиперпараметров.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - модели машинного обучения и анализа данных и их параметры; - понятие гиперпараметров, определение и настройка гиперпараметров; - принципы и подходы: поиск по решётке, случайный поиск, Байесовская оптимизация, оптимизация на основе градиентов, эволюционная оптимизация; - обзор инструментов для оптимизация гиперпараметров.
5	<p>Обучение моделей.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - разработка моделей анализа данных, машинного обучения; - Jupyter Notebook, ядра, Python, Google Colab; - среда разработки Spyder, DataSpell, Anaconda; - форматы моделей, хранилища моделей.
6	<p>Конвейер машинного обучения.</p> <p>Рассматриваемые вопросы:</p>

№ п/п	Тематика лекционных занятий / краткое содержание
	<ul style="list-style-type: none"> - понятие оркестрации моделей машинного обучения и анализа данных; - построение и автоматизация конвейеров машинного обучения; - MLOps - Machine Learning Operations; - обзор инструментов автоматизация конвейеров и оркестрации.
7	<p>Развертывание моделей.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - требования к эксплуатации моделей машинного обучения и анализа данных; - среды обучения и выполнения моделей машинного обучения и анализа данных; - варианты использования моделей, виды развертывания; - обзор инструментов развертывания.
8	<p>Интеграция моделей и мониторинг.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - особенности интеграции моделей машинного обучения и анализа данных в распределенных системах; - типовая архитектура сервисов для обучения и выполнения моделей машинного обучения и анализа данных; - проектирование и реализация прикладного интерфейса; - особенности мониторинга моделей машинного обучения и анализа данных; - метрики, оценка и показатели моделей; - логирование и анализ логов; - обзор инструментов мониторинга моделей.

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	<p>Постановка задачи и анализ данных.</p> <p>В результате выполнения лабораторной работы студент знакомится с особенностями постановки бизнес-требований к анализу данных, спецификациям и требованиям к эксплуатации разрабатываемой интеллектуальной системы.</p>
2	<p>Сбор, хранение и подготовка данных.</p> <p>В результате выполнения лабораторной работы студент получает навык работы с инструментами сбора, хранения и подготовки данных.</p>
3	<p>Контроль версий.</p> <p>В результате выполнения лабораторной работы студент получает навык версионирования данных и моделей с помощью инструментов Data Version Control и Liquibase.</p>
4	<p>Оптимизация гиперпараметров.</p> <p>В результате выполнения лабораторной работы студент получает навык оптимизации гиперпараметров моделей с помощью HyperOpt и Optuna.</p>
5	<p>Обучение моделей.</p> <p>В результате выполнения лабораторной работы студент получает навык работы с Jupyter Notebook, Spyder, DataSpell, Anaconda и Google Colab.</p>
6	<p>Конвейер машинного обучения.</p> <p>В результате выполнения лабораторной работы студент получает навык работы с Apache Airflow и Kedro для организации воспроизводимого конвейера машинного обучения.</p>

№ п/п	Наименование лабораторных работ / краткое содержание
7	Развертывание моделей. Python. В результате выполнения лабораторной работы студент получает навык работы с фреймворком Flask и FastAPI для реализации развертываемых сервисов моделей.
8	Развертывание моделей. Интеграция. Мониторинг. В результате выполнения лабораторной работы студент получает навык интеграции моделей в текущих цифровых интеллектуальных системах и развертывания средств мониторинга работы моделей.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение рекомендованной литературы.
2	Подготовка к лабораторным работам.
3	Подготовка к промежуточной аттестации.
4	Подготовка к текущему контролю.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Хапке, Х. Разработка конвейеров машинного обучения : руководство / Х. Хапке, К. Нельсон ; перевод с английского Н. Б. Желновой. — Москва : ДМК Пресс, 2021. — 346 с. — ISBN 978-5-97060-886-9. — Текст : электронный	https://e.lanbook.com/book/241088 (дата обращения: 10.04.2025)
2	Харенслак, Б. Apache Airflow и конвейеры обработки данных / Б. Харенслак, Р. Д. де ; перевод с английского Д. А. Беликова. — Москва : ДМК Пресс, 2022. — 502 с. — ISBN 978-5-97060-970-5. — Текст : электронный	https://e.lanbook.com/book/241133 (дата обращения: 10.04.2025)
3	Шарден, Б. Крупномасштабное машинное обучение вместе с Python : учебное пособие / Б. Шарден, Л. Массарон, А. Боскетти ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2018. — 358 с. — ISBN 978-5-97060-506-6. — Текст : электронный	https://e.lanbook.com/book/105836 (дата обращения: 10.04.2025)
4	Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения : руководство / С. Рашка ;	https://e.lanbook.com/book/100905 (дата обращения: 30.09.2025)

перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2017. — 418 с. — ISBN 978-5-97060-409-0	
---	--

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

Электронно-библиотечная система Научно-технической библиотеки РУТ (МИИТ) (<http://library.miit.ru/>)

Электронно-библиотечная система издательства «Лань» (<http://e.lanbook.com/>)

Открытые лекции (<https://sphere.vk.company/materials/video/#6>)

Открытые лекции (<https://sphere.vk.company/materials/video/#7>)

Открытые лекции (<https://sphere.vk.company/materials/video/#12>)

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

Браузер Microsoft Internet Explorer или его аналоги

Пакет офисных программ Microsoft Office или его аналоги

Python 3.6 и выше

Anaconda

Data Version Control

Liquibase

HyperOpt

Optuna

Spyder

DataSpell

Apache Airflow

MLflow

Kedro

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Учебные аудитории для проведения учебных занятий, оснащенные компьютерной техникой и наборами демонстрационного оборудования.

Для лабораторных работ – наличие персональных компьютеров вычислительного класса.

9. Форма промежуточной аттестации:

Зачет в 1 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

заведующий кафедрой, доцент, к.н.
кафедры «Цифровые технологии
управления транспортными
процессами»

В.Е. Нутович

старший преподаватель кафедры
«Цифровые технологии управления
транспортными процессами»

Е.А. Заманов

Согласовано:

Заведующий кафедрой ЦТУТП

В.Е. Нутович

Председатель учебно-методической
комиссии

Н.А. Андриянова